

Development of Possible Go-Around Criteria for Transport Aircraft

Angela M. Campbell*
Federal Aviation Administration
Atlantic City, NJ

Peter M.T. Zaal†
San Jose State University
NASA Ames Research Center
Moffett Field, CA

Jeffery A. Schroeder‡
Federal Aviation Administration
Moffett Field, CA

Somil R. Shah§
Federal Aviation Administration
Atlantic City, NJ

This paper adds data to help with the development of possible go-around criteria for transport category aircraft. Presently, airline procedures state that pilots make a go-around decision using multiple criteria at 1000 or 500 ft above the ground, or so-called gates. An experiment conducted on three level-D full-flight simulators investigated the conditions from which pilots cannot successfully recover from an unstable approach and make a normal landing. In addition, pilots' perceptions of risk under these various unstable approach conditions and resulting landings were assessed. Six crews, comprised of a captain and a first officer from the same airline, participated in each simulator. Both the captain and the first officer flew approaches and landings from 55 different initial conditions with varying gate heights, localizer deviations, glideslope deviations, reference-speed deviations, and rate of descents. The initial condition at the starting gate mainly affected longitudinal touchdown deviation and rate of descent at touchdown, with reference-speed deviation having the most significant effect. Results show little difference in touchdown performance for conditions from the 300-ft and 500-ft gates. Conditions at the 100-ft gate introduced significant differences in touchdown performance. Reference-speed and localizer deviation at the starting gate had the strongest influences on perceived risk and go-around decision. In line with other studies, these findings suggest that a 300-ft go-around gate might be acceptable. More research is required to investigate the effects of environmental and runway variables before possible go-around criteria for transport category aircraft can be defined.

Nomenclature

\dot{h}_{td}	sinkrate at touchdown, ft/s	GS	glideslope
n_r	number of runs with idle thrust, %	IATA	International Air Transport Association
V_{ref}	reference speed, kt	IMC	instrument meteorological conditions
x_{50}	runway distance left at 50 kts, ft	LOC	localizer
x_{td}	longitudinal touchdown point, ft	NASA	National Aeronautics and Space Administration
y_{td}	lateral touchdown point, ft	NTSB	National Transportation Safety Board
Abbreviations		PAPI	precision approach path indicator
		PFD	primary flight display
		RCC	runway condition code
		RWY	runway
		SFO	San Francisco International Airport
CAST	Commercial Aviation Safety Team	SME	subject matter expert
DOE	design of experiments	VMC	visual meteorological conditions
FAA	Federal Aviation Administration		
FSF	Flight Safety Foundation		

*Aerospace Engineer, William J. Hughes Technical Center, Atlantic City International Airport, NJ, 08405; angela.campbell@faa.gov. Member.

†Research Engineer, Human Systems Integration Division, Moffett Field, CA, 94035; peter.m.t.zaal@nasa.gov. Senior Member.

‡Chief Scientific and Technical Advisor for Flight Simulation Systems, Moffett Field, CA, 94035; jeffery.schroeder@faa.gov. Associate Fellow.

§Aerospace Engineer, William J. Hughes Technical Center, Atlantic City International Airport, NJ, 08405; somil.shah@faa.gov. Member.
The findings and conclusions in this paper are those of the author(s) and do not necessarily represent the views of the FAA.

I. Introduction

This paper adds data to answer the question whether today's stabilized approach criteria should be revised and detailed go-around criteria should be developed in order to reduce the number of approach and landing accidents and increase safety. Following the accident of American Airlines Flight 1420 at Little Rock National Airport in 1999, where the aircraft overran the runway following an unstable approach, the National Transportation Safety Board (NTSB) recommended that the Federal Aviation Administration (FAA) define detailed parameters for a stabilized approach, and develop detailed criteria indicating when a go-around should be performed.¹ In response to the NTSB recommendation, the FAA published stabilized approach guidance in advisory circular AC 91-79A that established targets for the approach.² Since the advisory circular does not provide specific guidance for when a go-around should be performed, the NTSB closed the recommendation in 2012 with an unsatisfactory response.

Since the accident in Little Rock, the go-around problem has gained international attention among airlines and industry safety groups.³ Most airlines have established policies that determine whether an approach is stable or unstable at specified altitudes or gates. Typically, airline procedures state that if a pilot determines the approach is unstable, they should go around. However, the collective industry performance of complying with go-around policies is extremely poor and only about 3% of unstable approaches result in a go-around. Studies by industry safety groups have suggested that the go-around noncompliance rate could be a significant safety hazard given that approach and landing are the most common phases of flight for aviation accidents, accounting annually for approximately 65% of all accidents.⁴ A Flight Safety Foundation (FSF) study of 16 years of runway excursions found that 83% could have been avoided with a decision to go around; thus, 54% of all accidents could potentially be prevented by going around.⁵

Improving the go-around compliance rate holds significant potential in reducing approach and landing accidents; consequently, organizations such as the Commercial Aviation Safety Team (CAST), the International Air Transport Association (IATA), and the FSF have investigated the root causes of go-around noncompliance and made recommendations for improving compliance rates.⁶ Studies by these organizations have revealed many reasons for flight crews to not follow go-around policies related to pilot judgment and company policies, including little-to-no consequence for not following the policies, lack of management awareness of the compliance rate, and pilot fatigue and situation awareness. One of the bigger factors is that pilots see current stabilized-approach criteria as too complex and restrictive for the operational environment.

Establishing an industry standard for missed approach criteria is challenging due to the large number of variables that influence approach and landing risk such as aircraft state, human factors, and environmental factors. Additionally, the criteria must mitigate the approach and landing risk without being so restrictive that they cannot be realistically implemented in today's operational environment. Recently, several airlines and safety organizations, including the FSF, have established revised stabilized approach criteria based on their approach and landing risk assessments. One example of revised criteria are the guidelines published by the FSF in a 2017 report which suggests that the go-around decision height can be lowered to 300 ft above ground level compared to 500 ft or 1,000 ft used by most airlines today.⁵

In response to the NTSB recommendation, unstable approach statistics, and industry trends, the FAA established a research program studying stabilized approach criteria. The overarching goal of the research is to determine whether it is feasible to develop universal, simplified stabilized approach criteria along with validating the recommendations from organizations such as the FSF. To address these challenges, a comprehensive research plan was developed, surveys were conducted, risk models were developed, and human-in-the-loop flight simulation experiments were planned. This paper details the first phase of experiments conducted during the fall of 2017 and adds to the literature as follows. First, it investigates the likelihood of an abnormal landing under various approach states encompassing current stabilized approach criteria. Second, it assesses pilots' perception of landing risk from these approach states by taking out the go-around decision-making process. Third, it used three level-D full flight simulators to compare results among narrow-body and wide-body aircraft, and among the Boeing and Airbus platforms. Finally, a sufficiently large pilot pool added statistical reliability to the results.

The paper is structured as follows. An overview of the research methodology is provided in Section II. The experiment setup is provided in Section III. The experimental results are provided and discussed in Sections IV and V, respectively, followed by the conclusions in Section VI.

II. Research Methodology

The research began with an extensive literature search and survey results to better understand current stable approach criteria and their shortcomings. This revealed that stable approach criteria definitions and compliance rates vary

widely throughout the industry. For many airlines, the go-around rate for unstable approaches is around 3% which indicates a mismatch between the industry's operational tolerance for unstable approaches and the written policies.

One of the more commonly cited reasons for not following stable approach criteria is that they are too restrictive and not realistic in operations. This sentiment was echoed in a 2015 survey conducted by the International Air Transport Association (IATA) in which pilots were asked to document their perceived shortcomings of stabilized approach criteria.⁷ Thus, the industry is facing a gap between the current policies and what is being practiced in operations, which leads to several questions: Why do pilots perceive go-around policies as restrictive? Are current go-around policies overly conservative? Are operators accepting too much risk in practice or a combination of both?

To help reduce the gap between go-around policies and practice, and to answer some of these questions, a human-in-the-loop experiment was designed with two primary objectives – 1) to determine under what conditions pilots cannot successfully recover an unstable approach and make a normal landing and 2) to assess pilots' perceptions of risk under various unstable approach conditions.

II.A. Task

The basic premise of the experiment was to correlate touchdown performance (dependent measures) with various approach states (independent variables) by requiring pilots to land the aircraft in a defined touchdown zone under a variety of initial-approach conditions in three level-D full flight simulators: B747-400, B737-800, and A330-200. Test subjects were required to always land the aircraft (no go-arounds allowed), even from conditions considered as an unstable approach by their airline and regardless of whether they personally felt that a go-around should be conducted. The purpose of this experiment design was to remove the go-around decision-making process to get an objective assessment of the likelihood of an abnormal landing under various approach states. The expectation was that under certain approach states (starting conditions), pilots would not be able to land in the specified touchdown zone or would have an excess speed or sink rate at touchdown. Under these conditions, the risk of an accident or incident would be elevated. Using these data, a determination can be made regarding the limits for which a pilot can land safely with an acceptable risk. These approach-parameter limits add data to help with the development of possible go-around criteria for transport aircraft.

All approach and landings were flown to runway 28R at San Francisco International Airport (SFO). Challenging environmental, runway, and aircraft parameters were selected to increase the difficulty of recovering from an unstable approach and landing the aircraft. The assumption was that go-around criteria developed under these extreme conditions would be conservative and applicable to more favorable conditions. All trials of the task were flown with a 10-knot tailwind, a wet runway with medium braking action, the maximum landing weight, and without any automation. Pilots were able to use localizer and glideslope error indicators on the primary flight display (PFD) and the precision approach path indicator (PAPI) at runway 28R at their own discretion. In addition, to time the flare of the aircraft, standard radio-altitude call-outs began at a main gear height of 50 feet and repeated in decrements of 10 ft until touchdown. Pilots were instructed to bring the aircraft to a full stop on the runway as quickly as possible. The flight card for the approach and landing task is provided in Fig. 1. The mean trajectory of one of the experimental conditions is provided in Fig. 2, along with the location of runways 28L and 28R, and taxiway C.

The approach and landings were flown by crews comprised of a captain and a first officer from the same airline. Both pilots alternated as the pilot flying and pilot monitoring. This allowed for objective data and subjective evaluations from both perspectives.

Task: approach and landing to SFO RWY 28R
Initial Condition: trim condition and location with respect to GS and LOC vary each run
Configuration: gear down, flaps full landing, speed brakes retracted
Weight: maximum landing weight
Ceiling/visibility: ceiling and visibility unlimited
Wind: 100/10 Turbulence: moderate Gusts: none
Runway: wet, medium braking action, RCC 3/3/3
Procedure:
1. Recover from a possible unstable approach to SFO RWY 28R (full recovery might not be possible)
2. Continue to land on RWY 28R
3. Flare and touchdown meeting, or as close to, desired touchdown criteria as possible
4. Apply thrust reversers and full manual braking
5. Task evaluation ends after the aircraft is fully stopped on the runway
Desired performance:
1. Longitudinal touchdown: 1,000 –2,000 ft from threshold
2. Lateral touchdown: centerline between main wing gear
3. Sink rate at touchdown: ≤ 6 ft/s
4. Bring the aircraft to a full stop as quickly as possible

Figure 1. Experiment flight card.

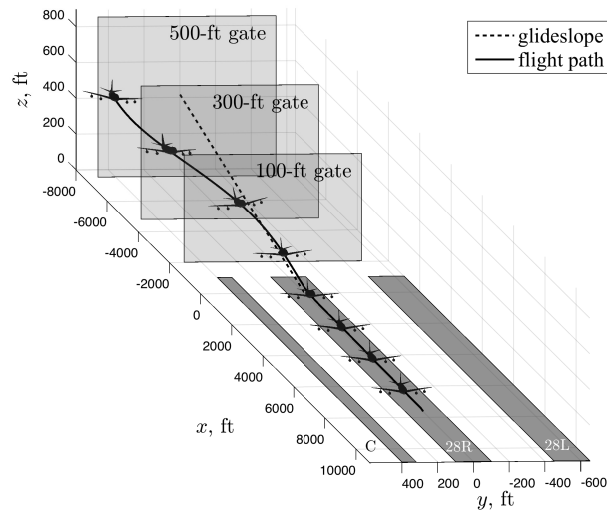


Figure 2. Mean B747 trajectory for the most extreme approach parameters.

Table 1. Typical stabilized approach criteria.

Parameter	Threshold
1 Stabilization height	1,000 ft IMC, 500 ft VMC
2 Localizer deviation	$\pm 1/2$ -1 dot
3 Glideslope deviation	± 1 dot
4 V_{ref} deviation	$V_{ref}-5 - V_{ref}+(10-20)$ kts
5 Rate of descent	1,000 ft/min
6 Power on approach	Above idle Appropriate with airspeed, approach, and condition
7 Aircraft configuration	Flaps in landing conf. Landing gear deployed Spoilers armed
8 Bank angle	Wings level 300-500 ft Less than 25 deg

II.B. Approach Parameters

The starting approach conditions were the independent variables for the experiment and are the parameters that would be used in the development of go-around criteria. The approach parameters for the experiment were selected based on information gathered during an extensive literature search, surveys of pilots, and an industry workshop. A study conducted by IATA revealed a total of eight stabilized approach parameters used by airlines (see Table 1);⁷ however, the surveys and workshop revealed five main parameters of interest: gate height, localizer deviation, glideslope deviation, V_{ref} deviation, and rate of descent. Other parameters, such as power on approach and bank angle, are less frequently found in stabilized approach criteria than the five parameters listed and were not included as independent variables in the current study. Nonetheless, they were tracked to determine whether they have any significant influence on touchdown performance.

The next step was to select the levels for each approach parameter to test in the experiment. The parameters were chosen such that some settings might be outside the range that a pilot can reasonably achieve a desired landing; however, the settings should not be so extreme that it would be a foregone conclusion that a successful landing would not be achieved. Therefore, the most extreme approach factor settings were selected to be slightly outside of typical stable approach criteria (Table 1).

The gate heights to test were selected first. Stabilization heights are typically 500 ft or 1,000 ft; however, some new guidelines published by the FSF suggest that the height could be lowered to 300 ft. Currently, it is generally accepted that the aircraft should be configured for landing by 1,000 ft; however, many industry subject matter experts (SMEs) that participated in the workshop suggested that the stabilization height for other parameters could be at 500 ft or below. Thus, the experiment considered three stabilization heights: 100 ft, 300 ft, and 500 ft (Fig. 2). The settings for the remainder of the approach parameters that were developed using the typical stabilized approach criteria in Table 1, SME input, and trials in a B747-400 Level-D simulator are provided in Table 2. During initial trials, it was found an unstable approach with a given set of stable approach criteria deviations might be recoverable from 500 ft, but would not be recoverable at 100 ft. For this reason, different levels for the approach parameters were chosen for each stabilization height.

To reduce the size of the final test matrix, only localizer deviations to the right of the runway were considered, assuming that deviations from the left and right would have similar but opposite effects. This also was more realistic with the parallel runway (28L) located to the left of runway 28R. In addition, only deviations above the glideslope were considered, as these are more difficult to compensate for than deviations below the glideslope. Furthermore, only speed deviations above the reference speed V_{ref} were tested; speed deviations below were considered a more dangerous situation that always warrants a go-around.

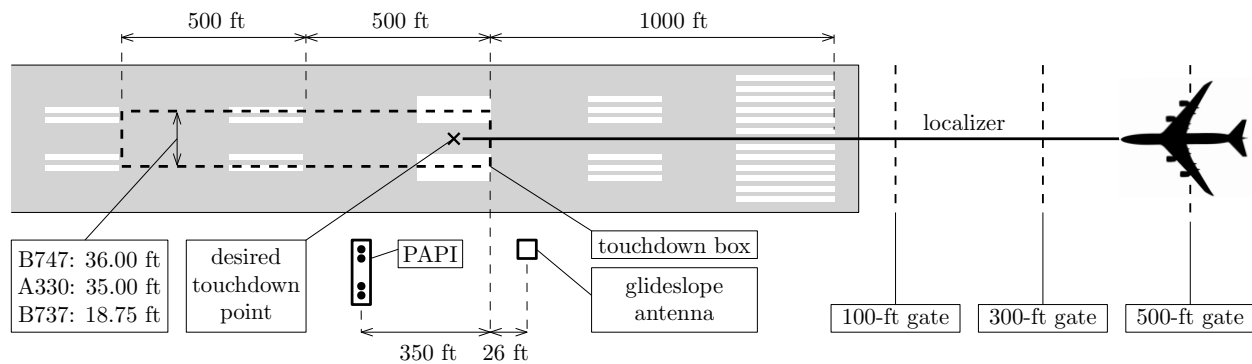


Figure 3. Touchdown zone definition.

II.C. Performance Criteria

The touchdown point was defined as where the center of the main gear touched the runway. Touchdown performance metrics were developed using SME input and published guidance. The selected metrics are representative of what would be considered a routine landing in normal operations. One metric of interest was the touchdown point. Landing too close to the threshold could lead to a higher risk of an undershoot, and landing too far down the runway could lead to a runway overrun. Additionally, large deviations from the centerline could lead to a runway veer-off. Based on these considerations, a touchdown box was created to bound the acceptable area for the aircraft to touchdown. The touchdown box began at 1,000 ft past the threshold and ended at 2,000 ft. The width of the box was equal to the width of the main gear of the aircraft, 36, 35, or 19 ft for the B747-400, A330-200, or B737-800, respectively. The touchdown box is illustrated in Fig. 3. The other metric of interest was rate of descent at touchdown. An unstable approach with high energy could potentially lead to a hard landing. Therefore, a threshold of 6 ft/s was used for the touchdown criteria.

II.D. Hypotheses

Based on the literature, pilot surveys, industry workshop, and test pilot trials in the B747-400 simulator, the following hypotheses were developed:

- H1: It was expected that initial approach conditions closer to the maximum values of the typical stabilized approach criteria in Table 1 would be more difficult to correct, resulting in landings outside of the touchdown performance criteria, and higher pilot workload and risk assessments. Furthermore, because of the nature of the aircraft energy-management problem, pilots would either land harder and short, or softer and long, following the more extreme initial approach parameters.
- H2: It was hypothesized that recovering from the extreme conditions at the 100-ft gate would be more difficult than recovering from extreme conditions at the 300-ft and 500-ft gates.
- H3: It was expected that reference speed deviations and higher rates of descent would be more difficult to recover from than localizer and glideslope deviations.

III. Experiment Setup

III.A. Independent Variables

The initial approach parameters were the independent variables of the experiment: stabilization height (100, 300, or 500 ft), glideslope deviation (0, 0.5, 0.75, or 1.5 dot), localizer deviation (0, 0.5, 0.75, or 1.5 dot), rate of descent (1000, 1250, or 1500 ft/min), and reference speed deviation (+0, +10, and +20 kts). The approach parameters, summarized in Table 2, had different levels at each gate height.

A test matrix was generated using the custom design-of-experiments (DOE) feature in JMP®⁸ to determine the number of runs and scenarios that each pilot would fly during the experiment. A full-factorial design was not used because the allotted time on the simulator would not allow for repeated conditions. The custom DOE option allowed replicates of the corners of the design while maintaining an experiment power of 1 for main effects and first order

Table 2. Independent variable settings.

Gate Height ft	Glideslope dot	Localizer dot	Rate of Descent ft/min	V_{ref} Deviation kts
100	0	0	1000/1250	+0/+10/+20
300	0/0.5	0/0.5	1000/1500	+0/+10/+20
500	0/0.75/1.5	0/0.75/1.5	1000/1500	+0/+10/+20

interactions. The test matrix had 55 different initial-approach-parameter conditions and a total of 84 conditions with replicates included.

The final test matrix had 92 runs, including 4 training runs and 4 extra conditions with varied environmental parameters. Data from the special conditions will be used to help develop the next phase of experiments that will evaluate the effect of environmental parameters on go-around criteria.

III.B. Apparatus

The experiment was carried out using three Level-D full flight simulators: a B747-400 simulator at NASA Ames Research Center (Fig. 4),⁹ and B737-800 and A330-200 simulators at the FAA Mike Monroney Aeronautical Center (Figs. 5 and 6). The three aircraft types tested provided the ability to compare results among narrow-body and wide-body aircraft, and among the Boeing and Airbus platforms. Using this approach allowed for a possible broadening of the criteria.

All simulators were from the same manufacturer and used in their standard configurations. Differences between simulators existed because of the different years of initial operation and the different aircraft types simulated. Care was taken to make all basic aircraft and environmental settings as similar as possible between simulators. For example, cockpit radio-altitude and warning call-outs, turbulence intensity, and runway and radio-navigation-aid geometries were equalized to provide pilots with similar basic cues across simulators. The out-the-window visual database was different between the A330/B737 and B747 simulators because of the different visual system manufacturers; however, basic features in the visual scene were equalized. Motion cues were provided in each simulator using the standard motion logic settings. The B747 and B737 simulators have a hydraulic hexapod motion system, and the A330 simulator uses an electric hexapod motion system.

All approach and landings were flown without any automation engaged. Autopilots, autothrottles, autobrakes, and flight directors were turned off. The PFD depicted conventional localizer and glideslope error indicators. Fig. 7 shows the PFD in the B737-800 simulator for the most extreme initial approach parameters (500 ft, 1.5 dot, 1.5 dot, 1500 ft/m, and $V_{ref}+20$ kts). The corresponding out-the-window visual is provided in Fig. 8. Note that in this condition, the aircraft is approximately lined up with the taxiway to the right of RWY 28R.

Pilots used tablet computers to fill out a questionnaire after each run. These tablet computers were mounted on the left and right sides of the simulator cabs for the captain and first officer, respectively, for easy access.

III.C. Controlled Variables

All the approaches were flown to SFO runway 28R. Runway 28R has a displaced threshold, and is 11870 ft long and 200 ft wide. The PAPI and instrument landing system glidepath are not coincident for this runway (Fig. 3).



Figure 4. B747-400 simulator.



Figure 5. B737-800 simulator.



Figure 6. A330-200 simulator.



Figure 7. B737-800 PFD.



Figure 8. Out-the-window view.

The environmental conditions were held constant for all runs because testing environmental parameters was not a goal of the experiment. The environmental conditions were selected to increase the difficulty of the landing. The assumption was that if the go-around criteria were developed under these conditions they would be applicable to more favorable conditions. All approaches were flown under VMC, and had a 10-knot tailwind, moderate turbulence, ceiling and visibility unlimited, and a wet runway with medium braking action (Fig. 1).

Additionally, the aircraft was at its maximum landing weight. Because the aircraft was always below 1,000 ft for the experiment, it was always in its landing configuration (gear down and flaps at the correct setting).

III.D. Participants

Six crews comprised of a captain and a first officer from the same airline participated in each simulator, giving a total of 18 crews or 36 pilots in the entire experiment. Crews from six different airlines participated. Most pilots were very experienced with ratings on many aircraft types. One B737 pilot was retired at the time of the experiment. All other A330 and B737 pilots were current and qualified as captain or first officer in a part 121 carrier. B747 pilots had to be flying for a part 121 carrier, and current and qualified as captain or first officer in the B747-400 in the last 36 months. All pilots gave written consent for their participation and received compensation.

III.E. Procedures

Each crew was scheduled for two consecutive days. Pilots were provided a briefing document, flight card (Fig. 1), airport diagram, and approach plate before the start of the experiment. On the first day, crews received an extensive pre-briefing, explaining the schedule, task, conditions, and procedures of the experiment. Crews were told the experiment investigated the effects of different approach parameters on landing performance and were given no specifics about the true nature of the experiment. Crews were informed they would fly 184 approach and landings with different initial conditions and that the conditions would be presented randomly. No further details were given about the extremity of the conditions. After the briefing, pilots provided their informed consent and filled out a pre-simulation questionnaire. This questionnaire gathered demographic data and information on their airline's current VMC stable approach criteria, and asked about their satisfaction with those criteria. The experiment started after a simulator safety briefing.

Each crew flew 92 scenarios per day for 2 days, divided among four 1-hour simulator sessions per day. Over the course of the 2 days of experimentation, each pilot flew all 92 scenarios in the test matrix. The runs were randomized for every pilot, and pilots rotated between the pilot-flying and pilot-monitoring role after each session. The first four runs of the first session were training runs with nominal approach conditions at the different gates.

The pilot flying flew each approach and landed the aircraft during every run of a session. The initial approach parameters were called out to the pilots by the experimenter before the start of the run to make sure pilots were completely aware of the situation. The pilot was asked to meet the touchdown criteria as closely as possible and then use maximum manual braking and full reverse thrust to bring the aircraft to a complete stop on the runway. The pilot monitoring was allowed to provide call-outs to assist the pilot flying, as per their airline policy or personal preference. After the aircraft had come to a complete stop, the simulator was reset for the next run, and the pilots answered their post-run questionnaires on the tablet computers (see Section III.F). Pilots were instructed to not discuss their post-run-questionnaire responses.

After completing all simulator runs, pilots filled out a post-simulation questionnaire. This questionnaire asked about pilots' preferred stable approach criteria based on what they experienced during the experiment and about the strongest influences on their decision to go-around. Finally, each crew received a debriefing providing more details about the true nature of the experiment.

III.F. Dependent Measures

Three main objective dependent measures specifying landing performance were recorded and analyzed: longitudinal and lateral touchdown location (x_{td} and y_{td}) and sinkrate at touchdown (\dot{h}_{td}). These measures related directly to the landing performance criteria pilots had to meet (Fig. 1). Data captures for these variables occurred when the main gear touched the runway. When multiple touchdowns were recorded (i.e., a bounced landing occurred), the maximum longitudinal distance and maximum sinkrate out of all touchdowns were used. The lateral distance always corresponded with the maximum longitudinal distance. In general, this meant that for a bounced landing, the sinkrate used belonged to the first touchdown, and the longitudinal and lateral touchdown location used belonged to the last touchdown.

In addition, two secondary objective dependent measures were evaluated: the number of runs with idle thrust in the approach (n_r) and the runway distance left at 50 kts (x_{50}). The thrust was determined idle in the approach when throttle inputs were zero at any time when the aircraft was higher than 50 ft above ground level. The runway distance left at 50 kts was calculated using the location on the runway when the speed of the aircraft was 50 kts.

Six subjective dependent measures were recorded using a questionnaire administered on a tablet computer at the end of each run. Pilots first rated their workload, fatigue, and the perceived risk of the previous landing (in that order) on a 20-point scale, by moving a slider bar with their fingers. Only the ends and midpoints of the slider bars were marked with low, average, and high. Next, pilots were asked if they would have called for a go-around based on personal risk criteria with a yes or no response. If they responded with yes, two questions followed. The first asked about the decision height for the go around and the second about the factors influencing their decision. Decision height was provided by giving a numeric input, and the factors were selected from a list. The list included the following factors: slow, fast, low descent rate, high descent rate, below glideslope, above glideslope, localizer deviation, power setting, and bank angle. Note, both the pilot flying and pilot monitoring filled out the post-run questionnaire, resulting in two sets of subjective data for each run.

IV. Results

At the completion of the experiment, data were collected from more than 3,000 approaches and landings from 36 pilots in total. The analysis focused on gathering statistical information on the aggregate data-set. Individual pilot or airline performance was not studied. The objective and subjective data were analyzed to determine the key factors that drive landing performance and landing decision making. The first step in the analysis was to construct a mixed model to determine the significant effects. Then, based on the results of the mixed-model analysis, the main effects of interest were explored further.

A mixed model is a statistical model that contains fixed and random effects. Fixed effects are those that are of primary interest to the study. Random effects have a large number of levels that influence the study results but are not of primary interest to the research.¹⁰ In this study, the fixed effects are the approach-variable and aircraft-type effects, and the random effects are subject effects. Mixed modeling analysis was chosen because it does not have the traditional analysis-of-variance-method restrictions on test design and balance of data.¹¹

Through the fit model platform in JMP®, a mixed model was created using the Restricted Maximum Likelihood (REML) estimation method.¹² Localizer deviation, glideslope deviation, reference speed deviation, rate of descent, and aircraft type were set as the fixed effects, and pilot ID was set as the random effect. The touchdown performance measures (longitudinal touchdown point, lateral touchdown point, and rate of descent at touchdown) were specified as the dependent variables. Using these inputs, JMP® produces the model parameter estimates, REML variance component estimates, and fixed effect tests.

The results of the fixed-effect tests are summarized in Tables 3 and 4. The first column of the table specifies the independent variables at each gate height. The mixed models were fit separately for each gate height because different levels of the independent variables were used at each gate. The second column states the degrees of freedom associated with the effect followed by the denominator degrees of freedom. The denominator degrees of freedom are based on the Kenward-Roger correction.¹³ The subsequent columns state the F-ratio and p-value for each independent-dependent variable combination. The p-value cells in the tables are shaded based on the JMP default conditional formatting to provide a visual representation of the significance of each effect.¹⁴ Dark shaded cells correspond to highly significant effects, light shaded cell represent significant effects, and white cells are non-significant.

Results presented below are aggregate data from all three simulators, except when otherwise noted. In that case, data are not linked to the specific aircraft types. Boxplots present medians and first and third quartiles of the aggregate data. The ends of the whiskers represent the lowest data point still within 1.5 times the interquartile range of the first

Table 3. Summary of main statistical test results for objective touchdown performance data.

Independent Variables		Dependent Measures					
		x_{td}		y_{td}		\dot{h}_{td}	
		df	F	p	F	p	F
500-ft Gate							
Localizer Deviation	2,1465	7.4494	0.0006	17.2039	<.0001	0.8494	0.4279
Glideslope Deviation	2,1465	10.3718	<.0001	0.3871	0.6791	5.1899	0.0057
Rate of Descent	2,1465	8.4898	0.0036	1.0643	0.3024	1.2071	0.2721
V_{ref} Deviation	2,1465	5.7672	0.0032	3.1521	0.0431	6.5151	0.0015
Aircraft Type	2,33	39.6204	<.0001	7.8227	0.0017	32.2118	<.0001
300-ft Gate							
Localizer Deviation	1,1035	0.0045	0.9464	33.0078	<.0001	0.1502	0.3984
Glideslope Deviation	1,1035	0.6012	0.4383	0.0000	0.9972	0.0117	0.9139
Rate of Descent	1,1035	0.7955	0.3727	0.2542	0.6143	0.1897	0.6633
V_{ref} Deviation	2,1035	9.8076	<.0001	1.1946	0.3033	5.7368	0.0033
Aircraft Type	2,32.98	36.1955	<.0001	2.6932	0.0825	28.3869	<.0001
100-ft Gate							
Rate of Descent	1,389.1	0.0005	0.9820	0.0005	0.9823	6.5829	0.0107
V_{ref} Deviation	2,389.2	54.8629	<.0001	1.9024	0.1506	12.1875	<.0001
Aircraft Type	2,1465	4.5181	0.0184	3.4391	0.0439	31.5357	<.0001







 = highly significant ($p < 0.01$)
 = significant ($0.01 \leq p < 0.05$)
 = not significant ($p \geq 0.05$)

Table 4. Summary of main statistical test results for workload and risk-rating data.

Independent Variables		Dependent Measures			
		Workload		Risk	
		<i>df</i>	<i>F</i>	<i>p</i>	<i>F</i>
500-ft Gate					
Localizer Deviation	1,3114	120.7110	<.0001	383.9251	<.0001
Glideslope Deviation	1,3114	56.1333	<.0001	206.3402	<.0001
Rate of Descent	1,3114	20.3099	<.0001	24.4788	<.0001
V_{ref} Deviation	1,3114	41.5081	<.0001	142.9267	<.0001
Aircraft Type	2,33	4.3562	0.0209	5.5005	0.0087
300-ft Gate					
Localizer Deviation	1,2106	7.6376	0.0058	14.5090	<.0001
Glideslope Deviation	1,2106	1.2062	0.2722	7.5575	0.0060
Rate of Descent	1,2106	7.8728	0.0051	39.1128	<.0001
V_{ref} Deviation	1,2106	11.6887	0.0006	116.4169	<.0001
Aircraft Type	2,33	3.8403	0.0317	5.8244	0.0068
100-ft Gate					
Rate of Descent	1,821	10.0089	0.0016	40.3569	<.0001
V_{ref} Deviation	1,821	21.2946	<.0001	90.6011	<.0001
Aircraft Type	2,33	4.2100	0.0235	8.0809	0.0014

 = highly significant ($p < 0.01$)
 = significant ($0.01 \leq p < 0.05$)
 = not significant ($p \geq 0.05$)

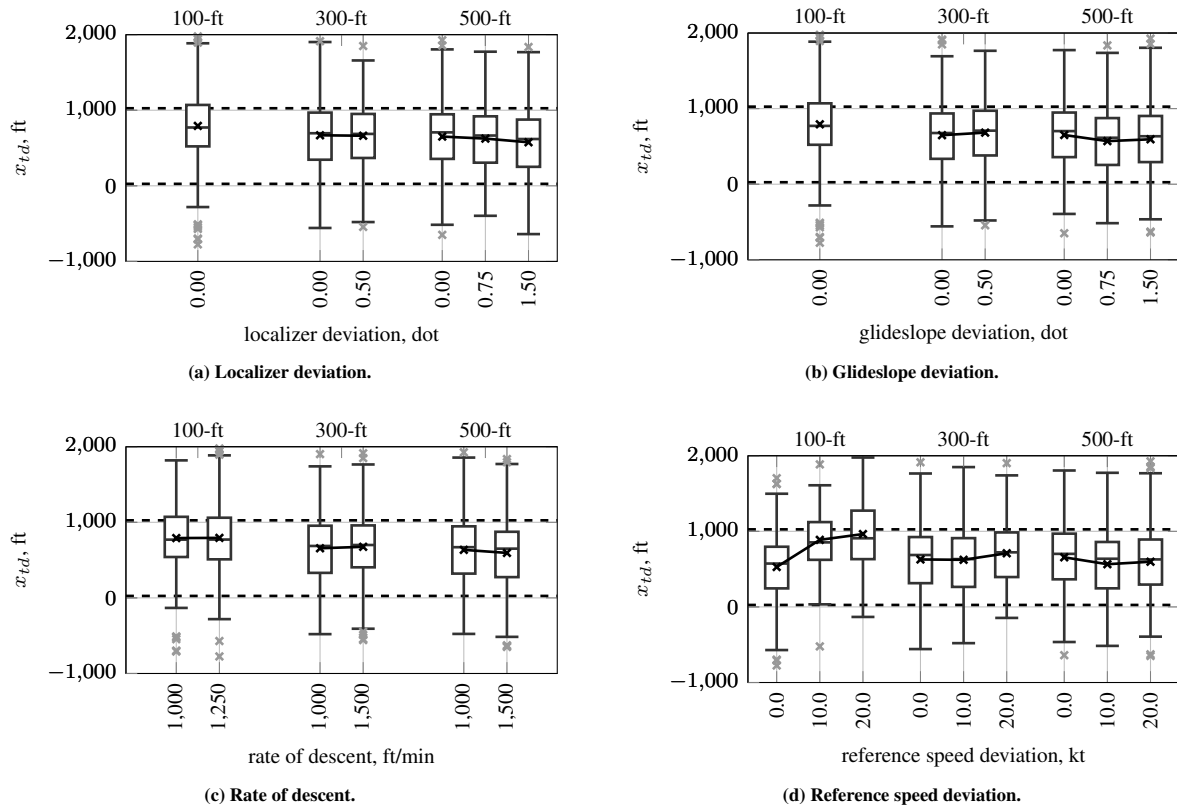


Figure 9. Effects of independent variables on longitudinal touchdown point.

quartile, and the highest data point still within 1.5 times the interquartile range of the third quartile. Data points outside of the whiskers are outliers and presented by gray x-symbols. The means of the data are depicted as well by black x-symbols. Dashed black lines indicate the touchdown performance criteria if applicable.

IV.A. Performance Data

The primary goal of the objective data analysis was to determine the effect of the approach parameters on touchdown performance. Along with the approach parameters, the effect of gate height and aircraft type were analyzed.

The touchdown performance metrics are depicted in Figures 9, 10, and 11. Data are grouped by decision height. Each figure contains four plots, one for each of the remaining independent variables (localizer deviation, glideslope deviation, rate of descent, and reference speed deviation). The results in Table 3 highlight the significant effects in the touchdown performance variables.

Fig. 9 shows the longitudinal touchdown point data. The effect of every independent variable on longitudinal touchdown point x_{td} was highly significant at the 500-ft gate height. Pilots touched down sooner with increasing localizer and glideslope deviations and higher rates of descent and reference speed deviations at the 500-ft gate. At the 300-ft gate, reference speed deviation and aircraft type (not shown) introduced highly significant differences in the longitudinal touchdown point between conditions. The longitudinal touchdown point increased for increasing reference speed deviations. Finally, reference speed deviation introduced highly significant differences and aircraft type significant differences in the longitudinal touchdown point between conditions at the 100-ft gate. A highly significant increase in the longitudinal touchdown point can be observed for increasing reference speed deviations (Fig. 9d).

Lateral touchdown point data are depicted in Fig. 10. It can be observed that more variability and outliers are present in the data for the 300-ft and 500-ft gates compared to the 100-ft gate. Few independent variables had a highly significant effect on lateral touchdown point. At the 500-ft gate, localizer deviation and aircraft type introduced highly significant differences and V_{ref} deviation significant differences in the lateral touchdown point y_{td} between conditions. Pilots touched down more to the right of the centerline with increasing localizer deviations to the right at the 500-ft gate. Pilots also touched down more to the right with increasing reference speed deviations. Localizer deviations at

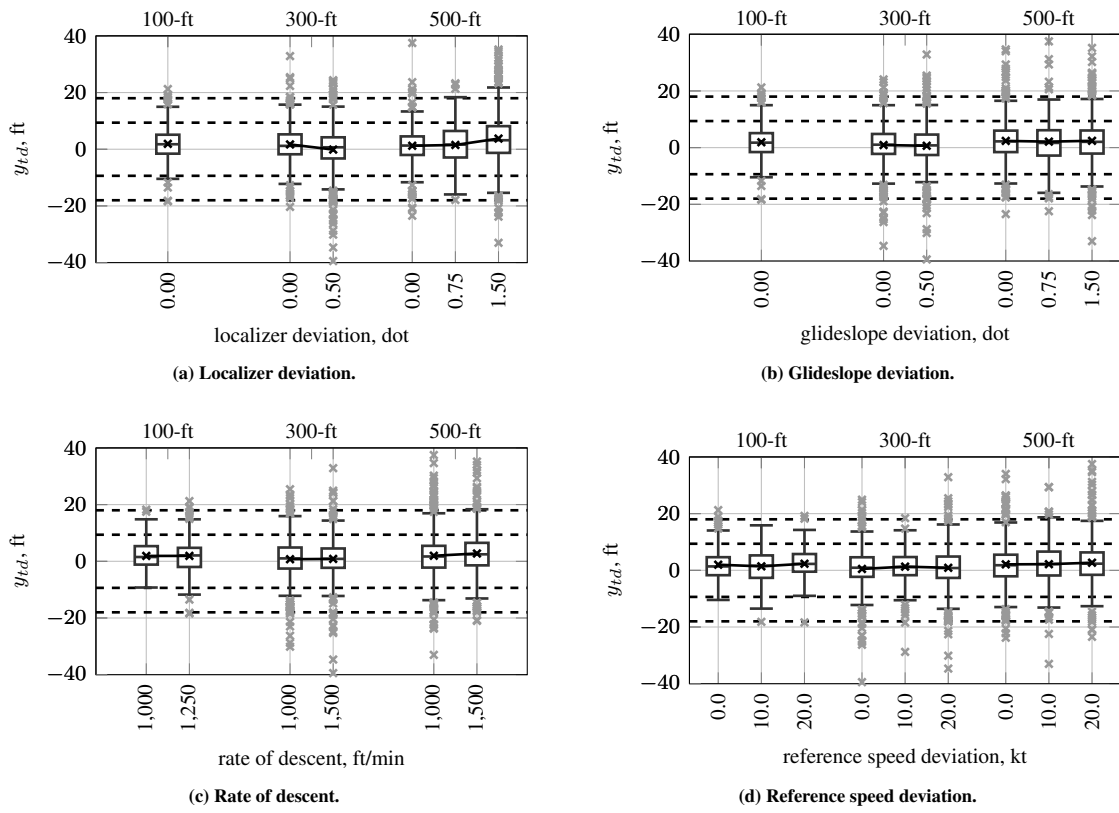


Figure 10. Effects of independent variables on lateral touchdown point.

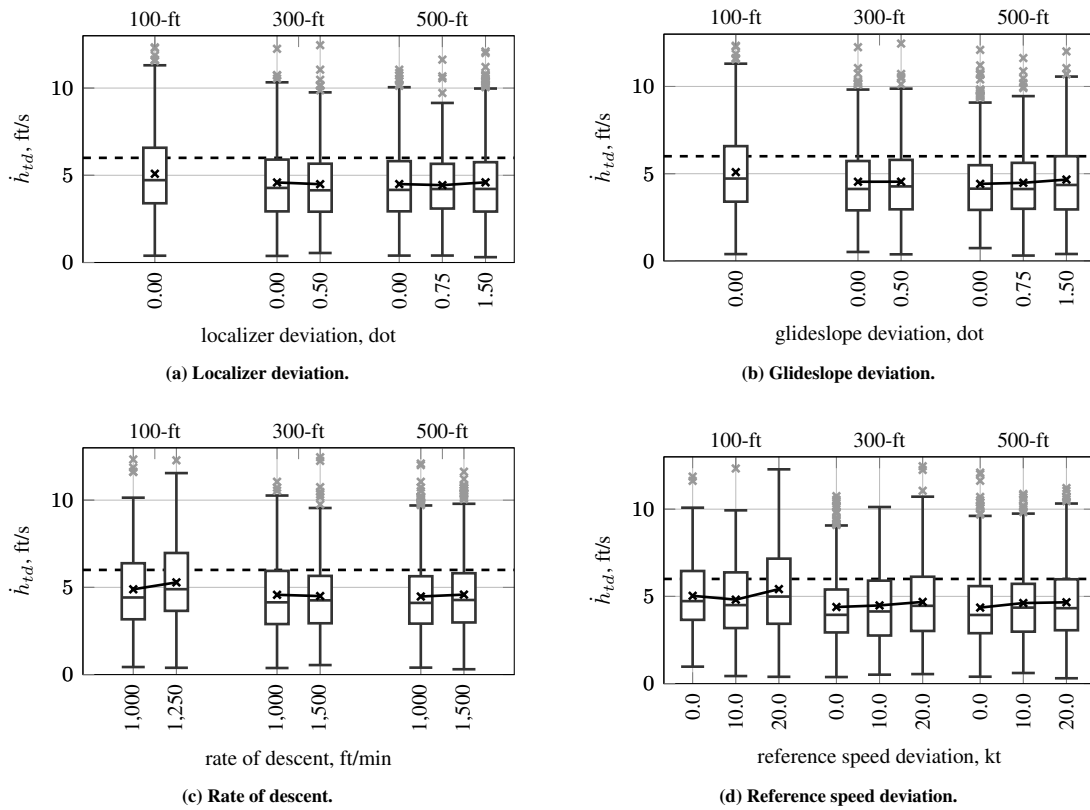


Figure 11. Effects of independent variables on sinkrate at touchdown.

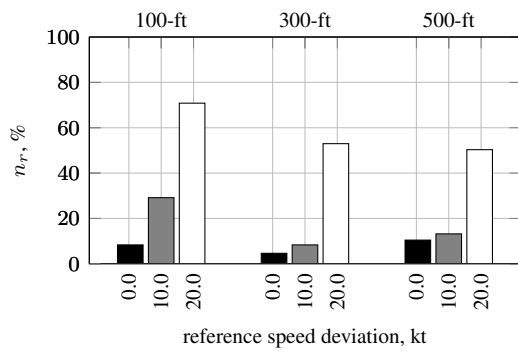


Figure 12. Effect of reference speed deviation on idle thrust in approach.

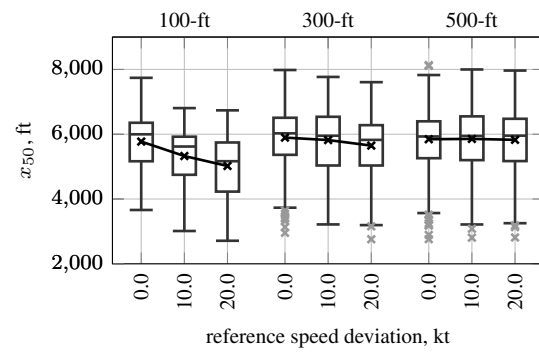


Figure 13. Effect of reference speed deviation on runway distance left at 50 kts.

the 300-ft gate introduced highly significant differences in the lateral touchdown point as well. Surprisingly, starting 0.5 dot to the right at 300-ft caused pilots to touch down more to the left compared to starting on the localizer at 300 ft. Finally, aircraft type introduced statistically significant differences in the lateral touchdown point at the 100-ft gate.

Fig. 11 depicts the sinkrate at touchdown. At the 500-ft gate, glideslope deviation, V_{ref} deviation, and aircraft type introduced highly significant differences in sinkrate at touchdown between conditions (Table 3). Starting increasingly higher above the glideslope at the 500-ft gate significantly increased the sinkrate at touchdown. In addition, starting with increasingly higher speed deviations above the reference speed significantly increased the touchdown sinkrate. Reference speed deviation and aircraft type introduced highly significant differences in sinkrate at touchdown between conditions at the 300-ft gate. Also starting with increasingly higher speed deviations at this gate significantly increased the sinkrate at touchdown (Fig. 11d). At the 100-ft gate, reference speed deviation and aircraft type introduced highly significant differences in the sinkrate at touchdown between conditions, and rate of descent deviations produced significant differences. For increasing rates of descent from 1,000 to 1,250 ft/min at the 100-ft gate, the sinkrate at touchdown marginally increased. In addition, the touchdown sinkrate increased for increasing V_{ref} deviations at the 100-ft gate.

Fig. 12 presents the percentage of runs with idle thrust in the approach grouped by gate height and reference speed deviation. The percentage of runs with idle thrust in the approach increased for increasing reference speed deviations at all gates. For $V_{ref} + 20$ kts, in more than 50% of the runs, pilots selected idle thrust during the approach. Starting further from the runway, that is, for increasing gate heights, the percentage of runs with idle thrust in the approach decreased. The largest differences in n_r between reference speed deviations were introduced at the 100-ft gate.

The runway distance left at 50 kts, grouped by gate height and reference speed deviation, is presented in Fig. 13. For increasing reference speed deviations, less runway distance was left at 50 kts. This effect was strongest for increasing reference speed deviations at the 100-ft gate and less strong for V_{ref} deviations at the 300-ft and 500-ft gates.

The mixed model analysis revealed that aircraft type had a strong effect on touchdown performance at all three gate heights, in particular on the longitudinal touchdown point and the sinkrate at touchdown. The plots in Fig. 14 provide more insights into the effect of aircraft type on these variables. Each x-symbol in Fig. 14 represents a landing. Pilots for aircraft type 3 had a tendency to land further from the threshold with a lower rate of descent. Conversely, pilots in simulator 1 landed with a higher rate of descent and closer to the desired touchdown point. The results for aircraft type 2 are in between those for types 1 and 3. The analysis brings into question the universality of go-around criteria.

Summarizing the performance data, primarily the fixed effects were significant on the longitudinal touchdown point and rate of descent at touchdown. Particularly, V_{ref} deviation and aircraft type had a highly significant effect on the dependent variables at all gate heights. The effect of V_{ref} deviation on the longitudinal touchdown point was highly significant or significant at all three gate heights, with the largest effects introduced for V_{ref} deviations at the 100-ft gate. Aircraft type had a highly significant effect on all the dependent variables except for the lateral touchdown point at the 300-ft gate.

IV.B. Subjective Data

Following every run in the simulator, both the pilot flying and the pilot monitoring completed a questionnaire. The questionnaire first asked pilots to rate their fatigue, workload, and landing risk for the previous run. Fatigue, workload,

and risk were rated on a scale between 0 and 20. A contour plot of risk evaluation for each run as a function of workload and fatigue is shown in Figure 15. The data around the edges of the contour plot are extrapolated from the ratings provided by pilots. The plot shows that the perception of landing risk for a run increased with workload and fatigue. In addition, workload had a stronger effect on perceived landing risk than fatigue.

After the risk question, the next screen of the questionnaire asked each pilot if he/she believed that a go-around should have been conducted on the previous run. If the pilot selected no, that was the end of the questionnaire. If the pilot answered yes, then he/she was asked at what height above ground level the go-around should have been conducted and why he/she would have elected to perform a go-around.

Figure 16 shows means and standard deviations of risk perception based on go-around decision. When pilots selected that he/she would have continued the landing, the mean risk score was about 7; if the pilot said a go-around should have been conducted, the mean risk score was about 13. The plot reveals that risk perception strongly influenced the go-around decision. When the landing risk was perceived to be higher, pilots more often answered that a go-around should have been conducted.

A mixed model analysis was performed to study the effect of the independent variables on the subjective dependent measures of workload and risk. A summary of the analysis can be found in Table 4. Except for glideslope deviation at the 300-ft gate, all the independent variables had a highly significant or significant effect on workload and risk scores.

Overall, localizer deviation and reference speed deviation had the strongest effects on the perception of risk (Table 4). Fig. 17 shows perceived landing risk grouped by gate height for varying localizer deviations and V_{ref} deviations. Fig. 17a shows that perceived risk increased significantly for increasing localizer deviations at the 300-ft and 500-ft gates. In addition, perceived risk was higher for the 100-ft gate conditions. Perceived landing risk increased for increasing reference speed deviations as well (see Fig. 17b). Particularly for the $V_{ref} + 20$ kts approach conditions, perceived landing risk was rated significantly higher.

V. Discussion

The data analysis revealed that touchdown performance for the 100-ft gate cases was significantly different than for the 300-ft and 500-ft gates. The results suggest that at the 100-ft gate, the aircraft was too close to touchdown for the pilots to correct the approach. Interestingly, the mean touchdown performance between the 300-ft and 500-ft gates did not vary significantly. The data indicate that the pilots were mostly able to correct the approach deviations at those gate heights.

Nevertheless, it is important to note that although the means of the touchdown parameters did not always vary much based on the starting conditions, the variance of the data was affected. The large deviations, such as the 1.5 dot localizer deviation at 500 ft, did tend to increase the variance of the touchdown parameters and produce more outliers. The presence of outliers should be taken into consideration when developing go-around criteria. The increased likelihood of an outlier could increase the risk of a runway excursion upon landing.

One goal of this study was to explore the universality of go-around criteria across aircraft types. Ultimately, the data showed that aircraft type significantly affected touchdown performance. The results shown in Fig. 14 revealed that aircraft type strongly affected whether the aircraft touched down soon on the runway with a high rate of descent

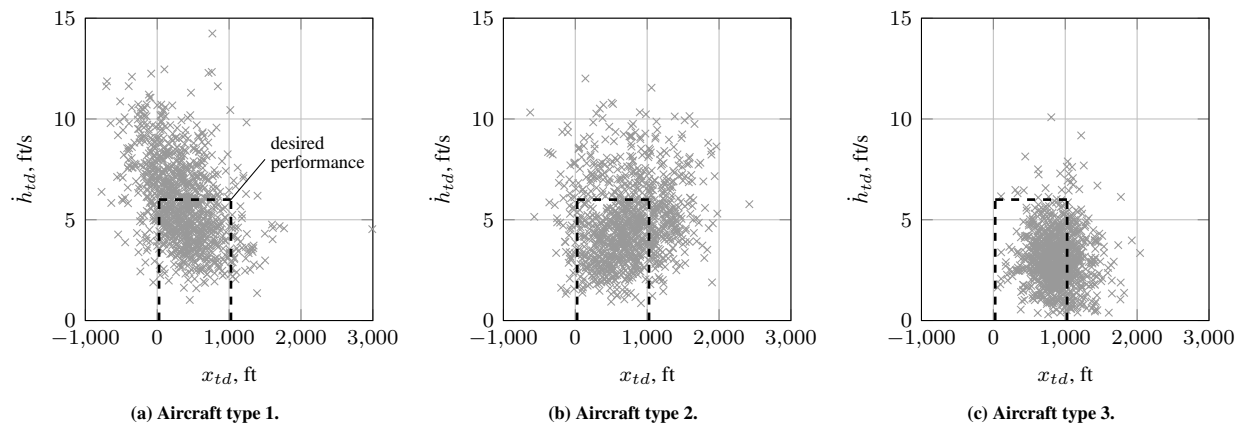


Figure 14. Rate of descent at touchdown vs. longitudinal touchdown point for each landing.

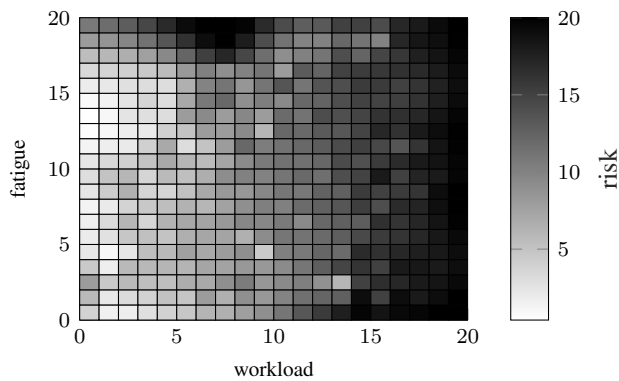


Figure 15. Risk perception vs. workload and fatigue.

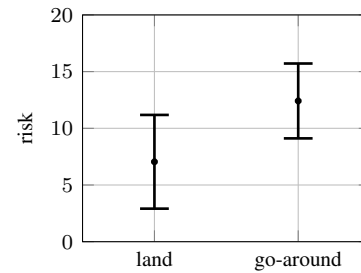


Figure 16. Risk perception vs. go-around decision.

or tended to touchdown further down the runway with a lower sink rate at touchdown. Although these data do reveal differences in the energy management techniques used by the pilots in the three aircraft types, further investigation is needed to determine whether the results suggest that the go-around criteria should be different for each aircraft type. Additional analysis showed that although the magnitudes of the touchdown parameters varied by aircraft type, the effects of the approach conditions were similar among all three aircraft types.

While not included as a variable in the experiment, thrust setting is a common parameter in stabilized approach and go-around criteria (Table 1). Often, the stabilized approach criteria threshold for thrust is above idle. Thus, the data were analyzed to determine which runs resulted in idle thrust use before crossing the threshold. The data revealed that idle thrust was often used when the V_{ref} deviation was 20 knots. Not surprisingly, many pilots set thrust to idle when given a V_{ref} deviation to reduce speed (often right at the start of the run). The data suggest that a low thrust setting is influenced by the speed of the approach. If the aircraft is on a normal approach profile, the likelihood that an inappropriate thrust setting is used would be reduced. To simplify the criteria, perhaps thrust setting could be eliminated; however, the thrust setting could be a good cross check to include in the criteria to ensure that the aircraft is at an appropriate energy state for the approach.

One purpose of collecting subjective questionnaire data was to determine whether pilots' perceptions of landing risk corresponded to the objective performance data. Although risk scores and go-around decisions were primarily a function of each individual pilot's risk tolerance, the initial approach parameters did have an effect. Approach states outside of typical stable approach criteria increased the workload scores for many pilots and affected their go-around decision making. Of the approach variables, V_{ref} deviation had one of the strongest effects on the pilot risk scores. The V_{ref} deviation also had one of the strongest effects on touchdown performance among the independent variables of the experiment, especially at the 100-ft gate.

Localizer deviation also significantly influenced risk scores. This likely occurred due to the jarring out the window view generated during the large localizer deviation runs. For the 1.5 dot localizer deviation, the aircraft was nearly lined up with the taxiway to the right of runway 28R (Fig. 8), thus triggering the high risk response from the pilots. Objectively, localizer deviations did not change mean touchdown performance significantly; however, as previously

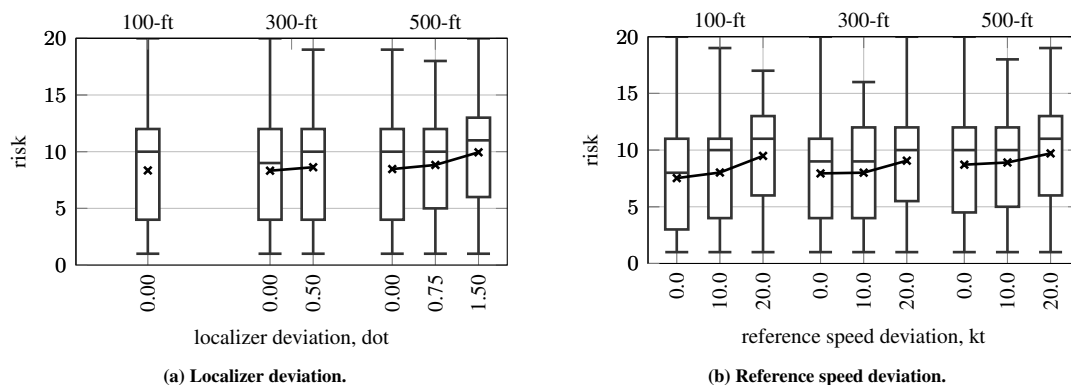


Figure 17. Effects of independent variables on risk perception.

discussed, large deviations did produce more outliers, which could increase risk. Mostly, the objective data and subjective scores were in agreement on approaches that could lead to unwanted outcomes. Thus, criteria developed on the objective data will likely coincide with pilot perception of risk.

All approaches were to runway 28R at SFO. The PAPI and ILS glidepath are not coincident for this runway; that is, when the localizer and glideslope error indicators on the PFD were zero, the PAPI could still indicate a deviation from the glideslope, depending on the distance from the runway. The severity of this mismatch depended on the aircraft type because of differences in the distance between the pilot station and the glideslope antenna. Pilots mainly relied on the instruments, so this most likely did not have an effect on the outcome of the experiment.

Finally, the simulator motion slowly ramped up at the start of each run. For the 300-ft and 500-ft gate conditions, there was enough time for the motion to ramp up before touchdown. However, because the 100-ft conditions started very close to the runway, the motion was not at full strength during touchdown, resulting in much more gentle touchdown bumps. In addition, this resulted in the turbulence being much less noticeable in the 100-ft conditions. This might possibly have been a confounding variable in the experiment that reduced the effects of the 100-ft starting conditions on landing performance.

VI. Conclusions

The purpose of this study was to investigate whether today's stabilized approach criteria should be revised and detailed go-around criteria should be developed. An experiment used three level-D full flight simulators with an approach and landing task under 55 different initial approach conditions to help the development of possible go-around criteria for transport aircraft. Eighteen crews participated, of which both the captain and first officer flew the entire test matrix for a perspective from both as the pilot flying or the pilot monitoring.

Both objective landing performance data and subjective workload and landing risk assessments were significantly affected by the initial approach condition. Touchdown performance was similar among the runs with a 300-ft and 500-ft starting gate. At these gates, pilots were often able to correct the deviations and meet the touchdown criteria. Conversely, the data from the 100-ft gate trials suggest that pilots did not have adequate time to correct any type of deviation, and their touchdown performance was significantly affected. These findings suggest that the go-around decision point should occur above the 100-ft gate, and a 300-ft go-around decision gate could be acceptable. At all gates, V_{ref} deviation had one of the stronger effects on touchdown performance. With increasing levels of fatigue and workload, the perception of landing risk increased. The reference speed and localizer deviations at the starting gate had the strongest and second strongest influence on the perceived risk and go-around decision, respectively.

The analysis presented in this paper focused on the main effects of the approach variables on the primary dependent measures. Additional analysis will be performed including investigating interaction effects and carrying out statistical tests on the secondary dependent measures, such as the use of idle thrust on approach. During this study, all environmental parameters were held constant. A follow-on study is required to determine whether the findings from this initial study hold true under different environmental conditions. The results of these studies intend to provide valuable information for the refinement of go-around criteria and development of go-around policies that will increase compliance while maintaining safety.

Acknowledgments

The authors would like to thank the simulation engineers and support personnel who contributed to the experiments. We especially thank Diane Carpenter and Richard Blanton from SimLabs at NASA Ames Research Center, and Randall Cooper, Ricky Zoellner, and Doug Rodzon from the Flight Operations Simulation Laboratory at the FAA Mike Monroney Aeronautical Center for their valuable contributions in setting up and running the experiments. We also thank John Hazelrig for being a test pilot on the B747 simulator. We would like to thank Hossein Eghbali from the FAA William J. Hughes Technical Center for the continued program support. Finally, we would like to thank the pilots who participated in the experiments and experiment-planning workshops.

References

- ¹National Transportation Safety Board, "Runway Overrun During Landing American Airlines Flight 1420 McDonnell Douglas MD-82, N215AA Little Rock, Arkansas June 1, 1999," Aircraft Accident Report NTSB/AAR-01/02, Oct. 2001.
- ²Federal Aviation Administration, "Runway Overrun Prevention," Advisory Circular 91-79A, Nov. 2007.
- ³Bureau d'Enquetes et d'Analyses, "Study on Aeroplane State Awareness during Go-around," Tech. rep., Aug. 2013.

- ⁴International Air Transport Association, “Unstable Approaches, Risk Mitigation Policies, Procedures and Best Practices,” Final report, 2016.
- ⁵Blajev, T. and Curtis, W., “Go-Around Decision-Making and Execution Project,” Final report, Flight Safety Foundation, March 2017.
- ⁶Commercial Aviation Safety Team, “Runway Excursion - Flight Crew Landing Training,” Safety Enhancement SE 216, June 2014.
- ⁷International Air Transport Association, “Unstable Approaches Survey,” Final report, 2015.
- ⁸JMP®, Version 13, Cary, NC: SAS Institute, 1989-2007.
- ⁹Sullivan, B. and Soukup, P., “The NASA 747-400 flight simulator - A national resource for aviation safety research,” *Flight Simulation Technologies Conference*, American Institute of Aeronautics and Astronautics, July 1996.
- ¹⁰SAS Institute Inc., “Mixed Models and Random Effect Models,” Tech. rep., Apr 2018.
- ¹¹Littell, R. C., “Analysis of unbalanced mixed model data: A case study comparison of ANOVA versus REML/GLS,” *Journal of Agricultural, Biological, and Environmental Statistics*, Vol. 7, No. 4, 2002, pp. 472–490.
- ¹²SAS Institute Inc., “JMP 10 Modeling and Multivariate Methods,” Tech. rep., 2012.
- ¹³Kenward, M. G. and Roger, J. H., “Small Sample Inference for Fixed Effects from Restricted Maximum Likelihood,” *Biometrics*, Vol. 53, No. 3, 1997, pp. 983.
- ¹⁴SAS Institute Inc., “Using JMP® 13,” SAS Institute Inc., Cary, NC, 2016.